nature research

Peer Review File

Manuscript Title: Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

This is a very clear, careful, and well-written analysis addressing a high-impact question. The authors find evidence of increased mortality associated with the new variant in the UK, even after adjusting for key confounders. The most convincing evidence to me is that they are able to look within local populations at concurrent cases with and without the variant, controlling for individual-level risk factors. A natural counter-explanation is that the more transmissible variant causes healthcare systems to be overwhelmed, leading to degradation in care and poorer outcomes, but this analysis provides important evidence that differences persist even beyond this.

This analysis will be of great interest to readers. The authors do a nice job stepping through each confounder, descriptive statistics, and modeling results, and presenting the evidence dispassionately and within the context of any limitations. Figures 1f-k are particularly helpful. I have only a few comments, primarily about some of the secondary analyses, though I do not expect changes to impact the final conclusions of the paper.

Major:
- In the misclassification analysis, what is the data source for estimating the false positive rate? I see that this has been embedded within a Bayesian framework with very wide priors, but intuitively, what is the data source here? How is this identified? Is there validation data internal or external to the study that might allow you to better estimate this quantity? Maybe the result is not sensitive here, but then if that is the case, that would be helpful to explain.

- Can the authors provide a reference for this strategy for correcting for misclassification error?

- For the absolute risk estimation, I understand that the authors chose something simple, but it seems inefficient since it only utilizes data from September 2020 plus the constant hazard ratio. In fact you have quite a bit of data to assess the case fatality ratios for later time periods, which could be generated from the model using something like a Breslow estimator (although the stratification could present a wrinkle). Worth considering, as it could improve precision.

Minor:
- Please add a reference for the Index of Multiple Deprivation

- As a comment, I do not view the missing hospital testing data as a weakness of the analysis, as these would be harder to interpret as they are along the causal pathway. Catching them as they emerge within the community seems best able to recreate prospective follow-up.

- As a comment, the discussion says that the analysis "could overestimate the change in mortality associated with SGTF if individuals infected by VOC die sooner than individuals infected with preexisting SARS-CoV-2 variants." But worth pointing out that the time-varying model that showed increased relative hazard at later times for SGTF vs. not.

Referee #2 (Remarks to the Author):

A. summary of key results to be updated by authors after revision made to exclude pillar 1 cases
B. Variant of Concern (VOC 202012/01) was identified in UK, is assessed as 50% to 70% more transmissible but impact on mortality undetermined until several research-teams submitted analyses to NERVAG, this being one (I assume). The current novel analysis of mortality of persons who tested PCR-positive in community in England [Pillar 2 tests via 3 named national laboratories] in exploits that S gene target failure [SGTF] is proxy for VOC.
C. Authors need to make clear if all Pillar 2 tests go via the 3 named national laboratories so that readers know if MISSING cases in Supplementary was missing because not-classifiable despite testing by one of the 3 laboratories OR are missing on account of either NOT-CLASSIFIABLE or tested at another Pillar 2 lab where classification is not feasible*. NOT-CLASSIFIABLE by one of the 3 named labs means Ct >=30 for ORF1ab or Ct >= 30 for N. Additional novelty is that authors allow for other variants than VOC to be falsely classified as SGTF at a constant low-level rate which on p11 is (I think) modelled regionally [despite no regional subscripts]. Hence authors replace binary SGTF (yes/no) with estimated pr (VOC). this nicety, in practice, make little difference in terms of inferences as shown in table 5.

C. & D. Data linkage is via a numerical identifier (please clarify WHICH). Linked are 3 data-files: i) specimen date and demography for test-subjects; ii) linelist of cycle threshold values fro ORF1ab, N and S; and iii) line-list of all deaths due to COVID 19 in England. Authors do not explain how line-list iii) is sourced, eg via Office for National Statistics (ONS) but coroner-referred deaths in England are not registered until coroner has determined cause of death (weeks or months later) + registration-delays . . . especially over festive season . . . are longer than usual so that allowance of only 10 days is somewhat meagre. Hospitals also report COVID-delays to dept Health/NHS England but again there are reporting delays. **

Authors fit proportional hazards (with/without allowance for non-proportionality, with geographical stratification or adjustment (based on NHS England region or Lower Tier Local Authority {LTLA . . . How many of these are there?? and does this really deal better with confounding]) with linear/spline age, linear/spline Index of Multiple Deprivation (IMD), with/without interactions. However, there are no likelihood ratios or other measures of fit to determine which models achieve parsimony + sufficiency without overfitting. in inferential terms, there is apparently little gained. BUT there is evidence that SGFT hazard-rate may greater at 14+ days post swab-date.

Figure 1 & 2 are rather small to read when printed out although of course work online with magnification.

Turning to corrected tables 1-3: i am surprised that authors analysed 1-34 years separately from 35-54 years [other than via linearity] as deaths are only 19/390K. in terms of IMD, rather than linearity, deprivation is generally significantly different between 1+2 vs 9+10 with 3-8 in between. The table I'd like to see has 120 cross-classifications as follows; specimen date (5) & age-group (4, starting at 1-54 years) * IMD (3, as above) * NHSE regions (2, ie the three regions where VOC was first noted versus the rest). For each cell, provide # SGTF cases, deaths, follow-up days and death-rates and similarly for non-SGTF cases.

Table 4 (deaths within any time period) could be relegated to Supplementary as inferences are not materially different and similarly Table 5 could be simplified for same reason; also because SGTF and pr (VOC) concur.

The table I'd bring out of Supplementary is the revised Table S1. As authors noted, overall 47% of PCR-positive cases are discarded from analysis because not classifiable wrt SGTF. HOWEVER, the missing-rate is strongly age-dependent (53% at 70-84 and 79% at 85+ years), likewise according

to place of residence (90% for care/nursing home), by NHS region (h55% to 63% for the 3 where SGTF first emerged) and by specimen date (around 35% fro 1st three periods but ? 50% for the most recent pair). These differences in %-missing may have been less evident before the Pillar 1 cases were removed but authors need to address why these differences have arisen and what implication s they have for their decision to analyse only the classifiable cases.*** The missing cases have very substantially higher death-rates!

Figures: Have Figures S1 etc been updated to remove Pillar 1 cases/ I ask because lablled as SGTF vs others, NOT as SGTF versus non-SGTF (as in other plots). Fig S2 shows non-proportionality at 70+ years, S8 shows limited follow-up for last specimen period.

Figures for 85+ years; authors do not mention vaccination against COVID19 that may affect 85+ year olds in their time-frame of analyses since vaccinations in UK began on 8 Dec.

E. CONCLUSIONS: wary of highly variable % missing and vulnerability to registration delay for deaths during and proximal to festive season. Extra 14 days of line-listing might help. Myriad models to deal with confounding - most lead to broadly similar inference bit paper does not help readers to discriminate parsimonious & sufficient from over the top!

F. IMPROVEMENTS; Additional 14 days of line-listing re deaths and consideration of implications of highly variable % missing for inferences that can be drawn from those classifiable. Why did authors not address more clearly the issues posed for highly variable % missing.

G. references sufficient.
H. Abstract to be re-drafted as consequence of Pillar 1 exclsuions.

Referee #3 (Remarks to the Author):

The authors perform an extensive analysis examining the association between VOC 202012/01 status and risk of death. Clearly a lot of work has gone into this and the authors are to be commended for it. My comments focus on the Abstract and Methods but don’t touch on the Results or Discussion. The reason for the latter is that I found how missing data in SGTF status was handled to be highly problematic. My comments below elaborate on this at a number of points in the manuscript but, crucially, that the rate of missingness is so much higher in the deaths than in the population writ-large is alarming. With that, I struggle to believe that a complete-case analysis is not subject to any bias, and so it is difficult to know what to make of the results. All this, however, can be resolved by the team and given the importance of the topic I would recommend that they be given an opportunity to do so.

Comments:

One thing that strikes me about the Abstract is that it doesn't indicate what the sample size was. This is particularly important when one eventually learns how much missing data regarding SGFT status there is.

L250: It is only in reading ahead and going through Tables S1 and S2 that one gets a sense of just how much missing data there is in SGFT status; indeed, the vast majority are missing. This seems highly problematic, especially given that the authors end up performing a complete-case analysis. Furthermore, having a sense of how much missing information there is seems incredibly important contextual information that should be in the main tables of the manuscript and not relegated to the supplement.

L253-6: Please clarify ... do Tables S1 and S2 represent all community tests and deaths in the UK during the study window? The captions don't make this clear. Beyond that, it seems highly

problematic that the missingness is so much greater among the deaths than the broader patient population. Later on the authors indicate the assumption that is necessary for the complete-case analysis to be valid but it seems that they could have done an investigation of the determinants of missingness and folded that in using, say, IPW.

L260: Some rationale for the 28-day window should be provided. I suspect it is because the authors want to sure the attribution of the death to the infection.

L261-2: An implication of this sentence is that the only form of censoring is that due to the fixed date of data extraction. If that is the case, then why not perform a logistic regression for the binary outcome of death 0/1 as the primary analysis? Is the actual timing of death (within the 28 day window) of interest?

L290-293: The authors make a case for the plausibility of the noninformative censoring assumption but not for the missing data assumption. As indicated in an earlier comment, much more needs to be done if we are to be convinced that missing data is not a big problem.

L301: Can the authors comment on the timing of the specimen relative to, say, when risk of death due to COVID actually occurs (i.e. from the timing of infection)? I suspect getting the latter is going to be very challenging but without any sense of the relation between the two dates its plausible that the results are subject to immortal time bias.

L305: This detail has already been indicated above, so could be removed.

L346: It may be that I am missing something but given that the dataset only includes ~7% of deaths and yet has ~47% of tests, won't any estimates of absolute risk be severely underestimated? Perhaps the authors are using data from the whole of the UK? This could be clarified.

L347: How is there any information in the data about risk of death at 60 days when person time was censored administratively at 28 days?

L347: As far as I can tell, "p_{VOC}" has not been specifically defined yet (at least not in the Methods). I assume it is actually the same as s(t) from the previous section but it's not clear.

L350: Since I imagine that there were secular trends over time in absolute risk, why not estimate "baseline risk" (which I am taking to be risk in the absence of SGFT) from the data during the study window using, say, an estimate of the baseline cumulative hazard function?

Author Rebuttals to Initial Comments:

Referee #1 (Remarks to the Author):

Note to all reviewers: We have now updated the analysis in response to reviewer comments, also incorporating an additional two weeks' worth of data. Note that the inclusion of additional data has revised our estimates for the increased hazard of death associated with SGTF upward.

This is a very clear, careful, and well-written analysis addressing a high-impact question. The authors find evidence of increased mortality associated with the new variant in the UK, even after adjusting for key confounders. The most convincing evidence to me is that they

are able to look within local populations at concurrent cases with and without the variant, controlling for individual-level risk factors. A natural counter-explanation is that the more transmissible variant causes healthcare systems to be overwhelmed, leading to degradation in care and poorer outcomes, but this analysis provides important evidence that differences persist even beyond this.

This analysis will be of great interest to readers. The authors do a nice job stepping through each confounder, descriptive statistics, and modeling results, and presenting the evidence dispassionately and within the context of any limitations. Figures 1f-k are particularly helpful. I have only a few comments, primarily about some of the secondary analyses, though I do not expect changes to impact the final conclusions of the paper.

Major:
- In the misclassification analysis, what is the data source for estimating the false positive rate? I see that this has been embedded within a Bayesian framework with very wide priors, but intuitively, what is the data source here? How is this identified? Is there validation data internal or external to the study that might allow you to better estimate this quantity? Maybe the result is not sensitive here, but then if that is the case, that would be helpful to explain.

We apologise for not including sufficient detail in the original manuscript. The data source is the relative frequency of Pillar 2 tests with SGTF over time by NHS England region, so it is not external to the study. However, when SGTF is adjusted for a baseline false positive rate—as we have done here—it corresponds closely with the frequency of VOC 202012/01 over time as identified by sequencing data (see “Update 1” from https://cmmid.github.io/topics/covid19/uk-novel-variant.html). We prefer to use SGTF data rather than sequencing data to estimate the frequency of VOC 202012/01 over time, since we know that the SGTF data are from the same population we are studying (whereas sequencing data come from a mix of community and hospital tests), and since there are many fewer sequenced samples available than samples with identified SGTF status.

However, to address the concerns of the reviewer, we have performed a simple sensitivity analysis calculating pVOC in a different way. Using COG-UK sequencing data from January 11 (the most recent data readily available to us), we count for each NHS region and date the number of samples that are confirmed VOC (i.e., lineage B.1.1.7 with deletion Δ69/Δ70 from Spike and mutation N501Y in Spike) = N(VOC), and the number of samples that are SGTF (i.e. samples of any lineage with deletion Δ69/Δ70, which causes S gene target failure) = N(SGTF), and estimate pVOC for a given NHS region and date as pVOC = N(VOC) / N(SGTF) from the sequencing data. If N(VOC) > N(SGTF) we set pVOC to 1. We also set pVOC to 1 for all dates after 31 December 2020, as the COG-UK data we have available does not include samples past January 1st. This sensitivity analysis, which uses sequenced SARS-CoV-2 samples from community and hospital tests in England rather than SGTF data from Pillar 2, yields very similar results to our main misclassification analysis (HR for pVOC is 1.63 [SE 0.07] in the main misclassification analysis, compared to 1.64 [SE 0.07] for this sensitivity analysis). We describe this additional sensitivity analysis in the Methods:

“The model above uses the same data source (i.e. SGTF frequencies among Pillar 2 community tests for SARS-CoV-2) as our survival analysis. To verify the robustness of this model, we performed a sensitivity analysis using sequencing data from the COVID-19 UK

Genomics Consortium[9] downloaded from the Microreact platform[10] on 11 January 2020 to estimate p_{VOC}. In this alternative analysis we estimated p_{VOC} for each NHS England region and date as the number of samples that were VOC 202012/01 (i.e. lineage B.1.1.7 with mutations ΔH69/ΔV70 and N501Y in Spike) divided by the number of samples that were SGTF (i.e. any lineage with ΔH69/ΔV70, the deletion that causes SGTF) for that NHS England region and date, setting p_{VOC} = 1 for all dates later than 31 December 2020 as there were no sequencing data available past this date, and filling any gaps in the data using linear interpolation. This yielded nearly identical results to our modelled probability of VOC (see **Fig. 2e**)."

- Can the authors provide a reference for this strategy for correcting for misclassification error?

This strategy is described by Keogh et al., 2020 (https://doi.org/10.1002/sim.8532) which we now cite in the "Misclassification analysis" section of the Results.

- For the absolute risk estimation, I understand that the authors chose something simple, but it seems inefficient since it only utilizes data from September 2020 plus the constant hazard ratio. In fact you have quite a bit of data to assess the case fatality ratios for later time periods, which could be generated from the model using something like a Breslow estimator (although the stratification could present a wrinkle). Worth considering, as it could improve precision.

While we agree that our method of estimating absolute risk is rather pragmatic, our aim in presenting absolute risks is to help communicate the meaning of the increased hazard (cf. Freeman and Spiegelhalter, https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1194-4). Since the baseline CFR for COVID-19 in England changes over time as new treatments are introduced (and potentially with novel variants), we feel that this simple approach fulfils the goal of communicating the magnitude of the increased hazard. We have rewritten this paragraph to clarify our objectives and to communicate the results with less spurious precision in the estimated risks:

"To put these results into context, we estimated how the absolute risk of death due to COVID-19 may change for an individual infected with VOC 202012/01. We calculated absolute risks by applying 28- and 60-day hazard ratios to the baseline (i.e. not variant-specific) risk of death estimated among individuals tested in the community between August–October 2020 (Table 3). The risk of death due to COVID-19 following a positive test in the community remains below 1% in most individuals younger than 70 years old. In females aged 70–84, the estimated risk of death within 28 days of a positive SARS-CoV-2 test with SGTF increases from 2.9% to 4.5% (95% CI 4.0–5.1%) and for females 85 or older increases from 13% to 20% (17–22%). For males aged 70–84 the risk of death within 28 days increases from 4.7% to 7.3% (6.4–8.2%) and for males 85 or older it increases from 17% to 26% (23–28%). Estimates of the absolute risk at 60 days were higher than those for 28 days and show similar patterns, with the largest increase of 20% to 30% (27–33%) for males 85 years and older. Estimates based on p_{VOC} were marginally higher. These estimates reflect a substantial increase in absolute risk amongst older age groups. Note that these estimates do not reflect the infection fatality ratio, but the fatality ratio among people tested

in the community, and are thus likely to be higher than the infection fatality rate as many infected individuals will not have been tested."

Minor:
- Please add a reference for the Index of Multiple Deprivation

We now cite the following technical report describing IMD at the first mention of IMD in the main text: https://www.gov.uk/government/publications/english-indices-of-deprivation-2019-technical-report

- As a comment, I do not view the missing hospital testing data as a weakness of the analysis, as these would be harder to interpret as they are along the causal pathway. Catching them as they emerge within the community seems best able to recreate prospective follow-up.

We thank the reviewer for pointing this out. We have added a sentence to the Discussion to reflect this: "However, by focusing on individuals tested in the community, our analysis captures any combined effect of an altered risk of hospitalisation given positive test and an altered risk of death given hospitalisation, which would not be fully captured by a study focusing on hospitalised patients only."

- As a comment, the discussion says that the analysis "could overestimate the change in mortality associated with SGTF if individuals infected by VOC die sooner than individuals infected with preexisting SARS-CoV-2 variants." But worth pointing out that the time-varying model that showed increased relative hazard at later times for SGTF vs. not.

We agree. After considering the issue further, we have removed this sentence as it seemed confusing to bring up this potential objection when it is not particularly evident in the analysis.

We thank the reviewer for these helpful comments.

Referee #2 (Remarks to the Author):

Note to all reviewers: We have now updated the analysis in response to reviewer comments, also incorporating an additional two weeks' worth of data. Note that the inclusion of additional data has revised our estimates for the increased hazard of death associated with SGTF upward.

A. summary of key results to be updated by authors after revision made to exclude pillar 1 cases

We have amended the summary accordingly.

B. Variant of Concern (VOC 202012/01) was identified in UK, is assessed as 50% to 70% more transmissible but impact on mortality undetermined until several research-teams submitted analyses to NERVAG, this being one (I assume). The current novel analysis of mortality of persons who tested PCR-positive in community in England [Pillar 2 tests via 3 named national laboratories] in exploits that S gene target failure [SGTF] is proxy for VOC.

C. Authors need to make clear if all Pillar 2 tests go via the 3 named national laboratories so that readers know if MISSING cases in Supplementary was missing because not-classifiable despite testing by one of the 3 laboratories OR are missing on account of either NOT-CLASSIFIABLE or tested at another Pillar 2 lab where classification is not feasible*. NOT-CLASSIFIABLE by one of the 3 named labs means Ct >=30 for ORF1ab or Ct >= 30 for N. Additional novelty is that authors allow for other variants than VOC to be falsely classified as SGTF at a constant low-level rate which on p11 is (I think) modelled regionally [despite no regional subscripts]. Hence authors replace binary SGTF (yes/no) with estimated pr (VOC). this nicety, in practice, make little difference in terms of inferences as shown in table 5.

We now give more detail on the reasons for missingness. In the main text, we specify "Of the 48% of tests with missing SGTF status, 9% were inconclusive due to high Ct values and the remaining 39% were not analysed in one of the three Lighthouse labs capable of producing an SGTF result."

C. & D. Data linkage is via a numerical identifier (please clarify WHICH).

We now state in the Methods that this is a numerical identifier for Pillar 2 tests called "FINALID". To our knowledge this identifier is specific to the Pillar 2 testing system.

Linked are 3 data-files: i) specimen date and demography for test-subjects; ii) linelist of cycle threshold values fro ORF1ab, N and S; and iii) line-list of all deaths due to COVID 19 in England. Authors do not explain how line-list iii) is sourced, eg via Office for National Statistics (ONS) but coroner-referred deaths in England are not registered until coroner has determined cause of death (weeks or months later) + registration-delays . . . especially over festive season . . . are longer than usual so that allowance of only 10 days is somewhat meagre. Hospitals also report COVID-delays to dept Health/NHS England but again there are reporting delays. **

We apologise for not giving more detail on the source of the deaths line list. The PHE COVID-19 deaths line list combines information from NHS England hospital deaths, Office for National Statistics death registrations, PHE Health Protection Team direct reporting, and Demographic Batch Service tracing of laboratory-confirmed cases (see https://www.gov.uk/government/publications/covid-19-reported-sars-cov-2-deaths-in-england/covid-19-confirmed-deaths-in-england-report for more details). We now cite this report and state these details in the Methods.

As for the allowance of 10 days being potentially insufficient, note that we include a sensitivity analysis restricting the analysis only to those individuals with a full 28 days of followup plus an additional 10 days (see Fig. 2e). This gives similar results to our main analyses.

Authors fit proportional hazards (with/without allowance for non-proportionality, with geographical stratification or adjustment (based on NHS England region or Lower Tier Local Authority {LTLA . . . How many of these are there?? and does this really deal better with confounding]) with linear/spline age, linear/spline Index of Multiple Deprivation (IMD), with/without interactions.

We now give the number of LTLAs (316), UTLAs (150), and NHS England regions (7) in the caption to Fig. 2.

Our main analysis uses LTLA + specimen date for stratification because we judged a priori that this finer stratification would be most capable of dealing with confounding at finer scales. One advantage of using stratification instead of adjustment for the finer geographical variables is that the models do not estimate one parameter per strata. We note this In the statistical methods section of the manuscript:

"The baseline hazard in the Cox model was stratified by both specimen date and LTLA, therefore finely controlling for these variables. The stratification gives a large number of strata matched by specimen date and LTLA. Only those strata that contain individuals who die and individuals who survive contribute to the analysis. The analysis is therefore similar to that which would be performed had we created a matched nested case-control sample."

We cannot test the assertion that this choice is better in controlling for confounding explicitly because our analysis does not independently estimate this source of confounding. However, we argue that in matched analysis it is well known that finer matches better control for confounders. Having said this, the choice of stratification does not have a substantial impact on estimated hazard ratios within the precision available to our analysis (Fig. 2e).

However, there are no likelihood ratios or other measures of fit to determine which models achieve parsimony + sufficiency without overfitting. in inferential terms, there is apparently little gained. BUT there is evidence that SGFT hazard-rate may greater at 14+ days post swab-date.

The model building is informed by likelihood ratio tests (for nested models using complete cases), or Wald tests (for the now implemented inverse-probability weighted models).
The manuscript methods section now reads:

"Models with and without interactions were compared using likelihood ratio tests for the complete cases analyses. For the analysis using IPW we used Wald tests based on robust standard errors"

To help guide the reader, we now focus on a smaller set of analyses in the first section of the results, and address the sensitivity analyses more briefly. We believe that the consistency of all analyses performed (Fig. 2e) suggests that overfitting is not generally an issue. While we do include a large number of sensitivity analyses, this is to demonstrate the robustness of our main analysis.

Figure 1 & 2 are rather small to read when printed out although of course work online with magnification.

We have enlarged the text of Figures 1 and 2 and reduced the number of panels in Figure 1. With the additional 14 days of data, the overall Kaplan-Meier plot no longer shows increased survival among SGTF subjects, as the confounding with age is now less extreme.

Turning to corrected tables 1-3: i am surprised that authors analysed 1-34 years separately from 35-54 years [other than via linearity] as deaths are only 19/390K.

We chose to include the 1-34 age group separately as there has been substantial interest in how the novel variant may impact upon younger individuals, particularly students of school and university age. While robust inference for an increased hazard among individuals younger than 35 is not possible given the small number of deaths in this group as the reviewer points out, we felt it was still of interest to show that rates of death remain very low in this age group.

in terms of IMD, rather than linearity, deprivation is generally significantly different between 1+2 vs 9+10 with 3-8 in between. The table I'd like to see has 120 cross-classifications as follows; specimen date (5) & age-group (4, starting at 1-54 years) * IMD (3, as above) * NHSE regions (2, ie the three regions where VOC was first noted versus the rest). For each cell, provide # SGTF cases, deaths, follow-up days and death-rates and similarly for non-SGTF cases.

This would certainly be possible, but we feel that the existing stratification within Figure 1, panels d–i, conveys the data in a format that is more accessible to readers than a table stratified by these 4 dimensions might be. We are happy to reconsider this if necessary.

Table 4 (deaths within any time period) could be relegated to Supplementary as inferences are not materially different and similarly Table 5 could be simplified for same reason; also because SGTF and pr (VOC) concur.

We have moved Table 4 to the SI and combined tables 1–3 to streamline the tables. We have also simplified Table 5 (now Table 3).

The table I'd bring out of Supplementary is the revised Table S1.

We now include missingness data originally in Table S1 as part of Table 1 in the main text.

As authors noted, overall 47% of PCR-positive cases are discarded from analysis because not classifiable wrt SGTF. HOWEVER, the missing-rate is strongly age-dependent (53% at 70-84 and 79% at 85+ years), likewise according to place of residence (90% for care/nursing home), by NHS region (h55% to 63% for the 3 where SGTF first emerged) and by specimen date (around 35% fro 1st three periods but ? 50% for the most recent pair). These differences in %-missing may have been less evident before the Pillar 1 cases were removed but authors need to address why these differences have arisen and what implication s they have for their decision to analyse only the classifiable cases.*** The missing cases have very substantially higher death-rates!

In order to address the missingness issue, we now carry out an inverse probability weighting (IPW) analysis which yields very similar results to the complete cases analysis. The missing cases have substantially higher death rates because there is an overrepresentation of care home residents among them, who are more likely to be tested with devices such as lateral flow antigen tests which cannot produce an SGTF reading. We now state this in the main text:

“SGTF status was missing in 89.1% of tests for individuals living in a care or nursing home, compared to 46.9% of tests among individuals in residential accommodation. This is partly due to more extensive use of lateral flow immunoassay tests in care homes, which do not yield an SGTF reading.”

Figures: Have Figures S1 etc been updated to remove Pillar 1 cases/ I ask because lablled as SGTF vs others, NOT as SGTF versus non-SGTF (as in other plots).

This was an inconsistency in labelling by us and we have now labelled all figures consistently as SGTF versus non-SGTF; these include only Pillar 2 tests with known SGTF status.

Fig S2 shows non-proportionality at 70+ years, S8 shows limited follow-up for last specimen period.

We test the proportional hazards assumption formally as part of our main analysis and indeed identify non-proportionality of hazards. We agree that there is limited follow-up for the last specimen period due to the date of data extraction; we have included a sensitivity analysis focusing only on individuals with complete 28-day followup to alleviate any concerns around this abbreviated follow-up period for more recent tests.

Figures for 85+ years; authors do not mention vaccination against COVID19 that may affect 85+ year olds in their time-frame of analyses since vaccinations in UK began on 8 Dec.

That is true, but our test for SGTF interactions with age does not turn up any significant differences in the hazard ratio by age. Since our analysis focuses on individuals who get tested (usually due to symptoms) and vaccine trials have suggested that the vaccines have similar efficacy against SGTF versus non-SGTF, it’s not clear that vaccination would have a

differential impact on mortality due to SGTF versus non-SGTF, and anything we could say about this would be fairly speculative.

E. CONCLUSIONS: wary of highly variable % missing and vulnerability to registration delay for deaths during and proximal to festive season. Extra 14 days of line-listing might help. Myriad models to deal with confounding - most lead to broadly similar inference bit paper does not help readers to discriminate parsimonious & sufficient from over the top!

We have streamlined which models we discuss more extensively in the main text, and moved discussion of other models to the “Further investigations” section to aid readers.

F. IMPROVEMENTS; Additional 14 days of line-listing re deaths and consideration of implications of highly variable % missing for inferences that can be drawn from those classifiable. Why did authors not address more clearly the issues posed for highly variable % missing.

There are now an additional 14 days of data in the analysis. Our IPW analysis addresses the issue of missing data and shows that our results are not due to patterns of missingness in the data.

G. references sufficient.
H. Abstract to be re-drafted as consequence of Pillar 1 exclsuions.

We have re-drafted the abstract to reflect the exclusions.

We thank the reviewer for these helpful comments.

Referee #3 (Remarks to the Author):

Note to all reviewers: We have now updated the analysis in response to reviewer comments, also incorporating an additional two weeks' worth of data. Note that the inclusion of additional data has revised our estimates for the increased hazard of death associated with SGTF upward.

The authors perform an extensive analysis examining the association between VOC 202012/01 status and risk of death. Clearly a lot of work has gone into this and the authors are to be commended for it. My comments focus on the Abstract and Methods but don't touch on the Results or Discussion. The reason for the latter is that I found how missing data in SGTF status was handled to be highly problematic. My comments below elaborate on this at a number of points in the manuscript but, crucially, that the rate of missingness is so much higher in the deaths than in the population writ-large is alarming. With that, I struggle to believe that a complete-case analysis is not subject to any bias, and so it is difficult to know what to make of the results. All this, however, can be resolved by the team and given the importance of the topic I would recommend that they be given an opportunity to do so.

To address the reviewer's concerns over the handling of missing data, we now perform an inverse probability weighting (IPW) analysis to correct for missingness in the data. It has not substantially changed our results, although the hazard ratios for the IPW analysis are slightly higher, with wider confidence intervals, than those for the complete cases analysis.

Further, we now explain in the manuscript that the missingness is much higher among care home residents than in the rest of the community:

"Having missing SGTF status was strongly associated with age and place of residence. The proportion with SGTF status missing was similar in age groups 1-34 (47.9%), 35-54 (47.1%) and 55-69 (47.7%), and then rose to 54.3% in the 70-84 age group and to 78.6% in the 85 and older age group. SGTF status was missing in 89.1% of tests for individuals living in a care or nursing home, compared to 46.9% of tests among individuals in residential accommodation. This is partly due to more extensive use of lateral flow immunoassay tests in care homes, which do not yield an SGTF reading."

This confounding with place of residence and age helps to explain why crude mortality is higher among individuals with missing SGTF status.

Finally, we apologise over the lack of clarity in the quoted figure from the previous version of the abstract of 7% of deaths due to COVID-19 in England being covered by the analysis. Our analysis looks at the population of individuals who were ever tested in the community. We have information on who died in this sample for all individuals (up to unreported deaths). The reason this only accounts for 7% of COVID-19 deaths in England is because most people with severe illness do not get tested in the community, but rather in the hospital.

Comments:

One thing that strikes me about the Abstract is that it doesn't indicate what the sample size was. This is particularly important when one eventually learns how much missing data regarding SGFT status there is.

We discuss the sample size in the first section of the Results, where we also now go into more detail about patterns of missingness in the data. This allows us to contextualise the sample size in terms of missingness and counts of SGTF versus non-SGTF samples.

L250: It is only in reading ahead and going through Tables S1 and S2 that one gets a sense of just how much missing data there is in SGFT status; indeed, the vast majority are missing. This seems highly problematic, especially given that the authors end up performing a complete-case analysis. Furthermore, having a sense of how much missing information there is seems incredibly important contextual information that should be in the main tables of the manuscript and not relegated to the supplement.

Just under half of community tests have missing data with respect to SGTF status. We now account for these missing data using an IPW analysis, and include the missingness in the data as part of Table 1 in the main text.

L253-6: Please clarify ... do Tables S1 and S2 represent all community tests and deaths in the UK during the study window? The captions don't make this clear. Beyond that, it seems highly problematic that the missingness is so much greater among the deaths than the broader patient population. Later on the authors indicate the assumption that is necessary for the complete-case analysis to be valid but it seems that they could have done an investigation of the determinants of missingness and folded that in using, say, IPW.

To clarify, there is no missingness in the deaths per se; however, tests such as lateral flow immunoassays which cannot assess SGTF status are more likely to be deployed in care homes, which means that there is a larger proportion of deaths among individuals with unknown SGTF status. As we now explain in the Results, “SGTF status was missing in 89.1% of tests for individuals living in a care or nursing home, compared to 46.9% of tests among individuals in residential accommodation. This is partly due to more extensive use of lateral flow immunoassay tests in care homes, which do not yield an SGTF reading.” While it is unfortunate that a large proportion of care home residents have missing SGTF status, we note that care home residents make up only a small fraction (<1%) of the general community, which is the target population of our analysis. Finally, we note that our new IPW analysis corrects for missingness in the data.

L260: Some rationale for the 28-day window should be provided. I suspect it is because the authors want to sure the attribution of the death to the infection.

The 28-day window is used because this is the standard way of counting deaths due to COVID-19 in England, which we now indicate in the main text. This is, in turn, indeed for reasons of attributing the death to the infection. We now include sensitivity analyses using a variety of different criteria for deaths, including those which have COVID-19 on the death certificate only.

L261-2: An implication of this sentence is that the only form of censoring is that due to the fixed date of data extraction. If that is the case, then why not perform a logistic regression for the binary outcome of death 0/1 as the primary analysis? Is the actual timing of death (within the 28 day window) of interest?

The timing of death within the 28-day window is of interest, as it may potentially illuminate differences in disease progression or testing behaviour among individuals with VOC 202012/01. The Cox analysis also allows us to use data from individuals without the full followup time, and to explore different lengths of follow-up without changing the form of the analysis.

L290-293: The authors make a case for the plausibility of the noninformative censoring assumption but not for the missing data assumption. As indicated in an earlier comment, much more needs to be done if we are to be convinced that missing data is not a big problem.

We have refitted all models using an IPW analysis, as previously explained (see Fig. 2e). This confirms that the patterns of missingness in the data do not have a substantial impact on our estimates of increased mortality.

L301: Can the authors comment on the timing of the specimen relative to, say, when risk of death due to COVID actually occurs (i.e. from the timing of infection)? I suspect getting the latter is going to be very challenging but without any sense of the relation between the two dates its plausible that the results are subject to immortal time bias.

Unfortunately these data are not available, but symptom onset generally occurs around 5 days after infection and death typically occurs around 3-4 weeks after infection. Given that our analysis now shows that test seeking behaviour seems relatively consistent between the new variant and preexisting variants (Fig. S12), we feel that immortal time bias is possible but unlikely.

L305: This detail has already been indicated above, so could be removed.

We have removed the earlier mention of the registration cutoff date.

L346: It may be that I am missing something but given that the dataset only includes ~7% of deaths and yet has ~47% of tests, won't any estimates of absolute risk be severely underestimated? Perhaps the authors are using data from the whole of the UK? This could be clarified.

Our analysis considers the population of individuals who get a test in the community. Among these individuals, the only missing death data would be due to non-reporting of deaths or potential misattribution of COVID-19 deaths to other causes, although the full data are supposed to include all deaths occurring within 60 days of a positive test. We now conduct several sensitivity analyses using different definitions of a COVID-19 death (Fig. 2e). The 7% deaths figure is the fraction of all COVID-19 deaths that occur in individuals who have been tested in the community. While this omits the larger proportion of COVID-19 deaths that occur among patients who are first tested upon hospital admission, by focusing on this group

we are able to capture an earlier stage in disease progression (infection -> symptoms -> test -> (hospital admission) -> death) and accordingly we obtain a fuller picture of changes to the risk of death. We also now compare the proportion of SGTF in Pillar 2 tests compared to randomly-sampled infections from the Office for National Statistics (Fig. S12) to show that a change in testing behaviour is unlikely to be driving the observed difference in the risk of death.

L347: How is there any information in the data about risk of death at 60 days when person time was censored administratively at 28 days?

Followup time was censored administratively at 28 days only for some of the survival analyses.

L347: As far as I can tell, "p_{VOC}" has not been specifically defined yet (at least not in the Methods). I assume it is actually the same as s(t) from the previous section but it's not clear.

We apologise for this omission and now define pVOC = s(t) / f(t).

L350: Since I imagine that there were secular trends over time in absolute risk, why not estimate "baseline risk" (which I am taking to be risk in the absence of SGFT) from the data during the study window using, say, an estimate of the baseline cumulative hazard function?

Our aim in presenting absolute risks is to help communicate the meaning of the increased hazard (cf. Freeman and Spiegelhalter, https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1194-4). Since the baseline CFR for COVID-19 in England changes over time as new treatments are introduced (and potentially with novel variants), we feel that this simple approach fulfils the goal of communicating the magnitude of the increased hazard. We have rewritten the paragraph presenting these results in the Results to clarify our objectives and to communicate the results with less spurious precision in the estimated risks:

“To put these results into context, we estimated how the absolute risk of death due to COVID-19 may change for an individual infected with VOC 202012/01. We calculated absolute risks by applying 28- and 60-day hazard ratios to the baseline (i.e. not variant-specific) risk of death estimated among individuals tested in the community between August–October 2020 (Table 3). The risk of death due to COVID-19 following a positive test in the community remains below 1% in most individuals younger than 70 years old. For the complete cases analysis, in females aged 70–84, the estimated risk of death within 28 days of a positive SARS-CoV-2 test with SGTF increases from 2.9% to 4.5% (95% CI 4.0–5.1%) and for females 85 or older increases from 13% to 20% (17–22%). For males aged 70–84 the risk of death within 28 days increases from 4.7% to 7.3% (6.4–8.2%) and for males 85 or older it increases from 17% to 26% (23–28%). Estimates based on IPW were marginally higher. These estimates reflect a substantial increase in absolute risk amongst older age groups. Note that these estimates do not reflect the infection fatality ratio, but the fatality ratio among people tested in the community, and are thus likely to be higher than the infection fatality rate as many infected individuals will not have been tested.”

We thank the reviewer for these helpful comments.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

I am satisfied with the authors' responses to my comments and those of the other reviewers. It is a strong paper, and readers will find it valuable.

Referee #2 (Remarks to the Author):

The paper is greatly strengthened by the additional follow-up, sharpened presentation of salient analyses, revised tables and by the cauchit Inverse probability weighting to account for missingness [please add reference re cauchit model as less familiar than logistic etc].

Thank you for the added explanations re contributing laboratories and also critical LFT explanation re high rate of missingness wrt older citizens/those in residential/nursing care - which is also a warning to UK about "undetected" vulnerability to new variants in the face of insufficient PCR testing.

The nett effect is that the paper's re-estimated HR, whether at 1.6 (95% CI 1.4 to 1.8) without IPW [but note time-varying HRs] and 1.7 (1.46 to 1.90 with IPW [time-varying insufficiently supported], is substantially higher than earlier hostages to fortune voiced pre-sufficient-analyses in UK. Authors might add a cautionary note along these lines to their Discussion; ie better the right answer a little later than early-wrong steer before detailed analysis could be undertaken. Notably and troublingly, this level of HR countermands the effectiveness of treatments for COVID-19 discovered in 2020.

In Discussion, authors might also add a note of regret (by referee, if not by them) that vaccination status could not be taken into account. I do not buy the authors' counter-argument re absence of interaction with age 85+ years because vaccine could reduce HR (conditional on being a case . . . although this is not proven as yet from RCT where hospitalization-rate per 100 cases is not significantly different according to randomized allocation) while VOC enhances HR: the nett effect, as measured here, is not thrilling to contemplate.

FINALID is now described as an/the identifier common to all three data-sources but samples do not arrive into labs with "FINALIDs" and so there must be a back-office mapping exercise - more or less fuzzy - from personal identifying information at lab or in association with deaths (2 sources) onto FINALID. Hence, what % of SGTF-differentiable PCR positives at the 3 labs capable of differentiation cannot be assigned FINALID? Presumably that are counted in the 3 labs' on-differentiables?

Re 1-34 years: I appreciate that authors have included due to public health concern, but it is equally important that their answer to referee [re absence of robust evidence about HR at 1-34 years] is also conveyed to readers . . .

In Fig 1, I note with interest that IMD gradient flattened with age . . . I don't know whether this stands up in multifactorial analysis (via interaction) & whether it does or not is NOT a main issue for this analysis BUT so little does flatten that IMD gradient that it could be of interest as a brief communication elsewhere!

Finally, I'd still like to see in supplementary material the several-way cross-tabulation that I asked

for in my original report as the re-analysis confirms those characteristics as key.

Referee #3 (Remarks to the Author):

On the whole the authors have been responsive, but I do have a number of remaining comments/concerns:

Abstract: In my original review I had asked for the authors to include some information on sample size. The authors have (implicitly) declined, suggesting that such information would need to be contextualized. While I understand their point, it seems to me that this information (in particular the missingness in the SGTF status) remains critical to understanding the paper. Moreover, I feel that it should not be incumbent on the reader to have to go through the manuscript to find this out.

L79-92: This paragraph provides a summary of missingness in SGTF status. What does not come across here, however, is that (from Table 2) ~74% (9,408/12790) of deaths have missing SGTF status. Given that the overall missingness (from Table 1) is ~48%, it seems clear that there is differential missingness by the outcome status. This is a point I tried to make in my initial review but, seemingly, didn't not come across as well as I would have hoped. Some mention/discussion of this should be given. It may be that the authors argue conditional independence between missingness in the SGTF status and outcome, specifically given the things that they include in the model that underpins the IPW analyses. If so, I would ask that they are explicit about that.

Table 1 seems to have two columns that provide the same counts (although the percentages have different denominators). I'd suggest combining in some way.

Table 2. Although a reader could work it out, percentages regarding missingness should be given. For example, the fact that ~74% of deaths have missing SGTF status (9,408/12790) is not evident.

Author Rebuttals to First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

I am satisfied with the authors' responses to my comments and those of the other reviewers. It is a strong paper, and readers will find it valuable.

We thank the reviewer for this assessment.

Referee #2 (Remarks to the Author):

The paper is greatly strengthened by the additional follow-up, sharpened presentation of salient analyses, revised tables and by the cauchit Inverse probability weighting to account for missingness [please add reference re cauchit model as less familiar than logistic etc].

We have added a reference to a review paper by Seaman & White which discusses the use of robust regression.

Thank you for the added explanations re contributing laboratories and also critical LFT explanation re high rate of missingness wrt older citizens/those in residential/nursing care - which is also a warning to UK about "undetected" vulnerability to new variants in the face of insufficient PCR testing.

The nett effect is that the paper's re-estimated HR, whether at 1.6 (95% CI 1.4 to 1.8) without IPW [but note time-varying HRs] and 1.7 (1.46 to 1.90 with IPW [time-varying insufficiently supported], is substantially higher than earlier hostages to fortune voiced pre-sufficient-analyses in UK. Authors might add a cautionary note along these lines to their Discussion; ie better the right answer a little later than early-wrong steer before detailed analysis could be undertaken. Notably and troublingly, this level of HR countermands the effectiveness of treatments for COVID-19 discovered in 2020.

We now mention this point in the Discussion: “Our findings here are consistent with earlier informal reports by ourselves and other groups assessing the risk of death among individuals with SGTF, although the increase in mortality reported here is higher than we had previously identified[8].” We feel it is not for us to comment here on the timing of the UK government’s announcement regarding increased mortality among SGTF cases.

In Discussion, authors might also add a note of regret (by referee, if not by them) that vaccination status could not be taken into account. I do not buy the authors' counter-argument re absence of interaction with age 85+ years because vaccine could reduce HR (conditional on being a case . . . although this is not proven as yet from RCT where hospitalization-rate per 100 cases is not significantly different according to randomized allocation) while VOC enhances HR: the nett effect, as measured here, is not thrilling to contemplate.

We have added a note on this limitation to the Discussion: “We were unable to account for vaccination status in this analysis.”

FINALID is now described as an/the identifier common to all three data-sources but samples do not arrive into labs with "FINALIDs" and so there must be a back-office mapping exercise - more or less fuzzy - from personal identifying information at lab or in association with deaths (2 sources) onto FINALID. Hence, what % of SGTF-differentiable PCR positives at the 3 labs capable of differentiation cannot be assigned FINALID? Presumably that are counted in the 3 labs' on-differentiables?

Unfortunately we do not have access to that information, but NHS number is routinely collected as part of community testing and this is likely to form a part of assigning a FINALID.

Re 1-34 years: I appreciate that authors have included due to public health concern, but it is equally important that their answer to referee [re absence of robust evidence about HR at 1-34 years] is also conveyed to readers . . .

We apologise for omitting this from the manuscript. We now state in the results: “We note, however, that the relatively small number of deaths among 1–34-year-olds over the study

period (44 deaths) does not permit robust assessment of the impact of SGTF in this age group."

In Fig 1, I note with interest that IMD gradient flattened with age . . . I don't know whether this stands up in multifactorial analysis (via interaction) & whether it does or not is NOT a main issue for this analysis BUT so little does flatten that IMD gradient that it could be of interest as a brief communication elsewhere!

We thank the reviewer for pointing this out and agree that it may be of interest for further analysis.

Finally, I'd still like to see in supplementary material the several-way cross-tabulation that I asked for in my original report as the re-analysis confirms those characteristics as key.

We have included this cross-tabulation as Supplementary Table 2.

We thank the reviewer for these helpful comments.

Referee #3 (Remarks to the Author):

On the whole the authors have been responsive, but I do have a number of remaining comments/concerns:

Abstract: In my original review I had asked for the authors to include some information on sample size. The authors have (implicitly) declined, suggesting that such information would need to be contextualized. While I understand their point, it seems to me that this information (in particular the missingness in the SGTF status) remains critical to understanding the paper. Moreover, I feel that it should not be incumbent on the reader to have to go through the manuscript to find this out.

We apologise for not including this information in the abstract previously. As requested, we now include the sample size of tests and of deaths with missing versus non-missing status in the abstract. The new abstract is:

SARS-CoV-2 lineage B.1.1.7, a variant first detected in the United Kingdom in September 2020, has spread to multiple countries worldwide. Several studies have established that B.1.1.7 is more transmissible than preexisting variants, but have not identified whether it leads to any change in disease severity. We analyse a dataset linking 2,245,263 SARS-CoV-2 community tests and 17,452 COVID-19 deaths in England from 1 September 2020 to 14 February 2021. For 1,146,534 (51%) of these tests, the presence or absence of B.1.1.7 can be identified because of mutations in this lineage preventing PCR amplification of the spike gene target (S gene target failure, SGTF). Based on 4,945 deaths with known SGTF status, we estimate that the hazard of death associated with SGTF is 55% (95% CI 39–72%) higher after adjustment for age, sex, ethnicity, deprivation, care home residence, local authority of residence and test date. This corresponds to the absolute risk of death for a 55–69-year-old male increasing from 0.6% to 0.9% (95% CI 0.8–1.0%) within 28 days after

a positive test in the community. Correcting for misclassification of SGTF and missingness in SGTF status, we estimate a 71% (48–97%) higher hazard of death associated with B.1.1.7. Our analysis suggests that B.1.1.7 is not only more transmissible than preexisting SARS-CoV-2 variants, but may also cause more severe illness.

L79-92: This paragraph provides a summary of missingness in SGTF status. What does not come across here, however, is that (from Table 2) ~74% (9,408/12790) of deaths have missing SGTF status. Given that the overall missingness (from Table 1) is ~48%, it seems clear that there is differential missingness by the outcome status. This is a point I tried to make in my initial review but, seemingly, didn't not come across as well as I would have hoped. Some mention/discussion of this should be given. It may be that the authors argue conditional independence between missingness in the SGTF status and outcome, specifically given the things that they include in the model that underpins the IPW analyses. If so, I would ask that they are explicit about that.

We apologise for the lack of clarity here. We do explicitly describe this assumption in the Methods:

"We performed complete cases analysis, restricted to the subset with SGTF status measured. This complete case analysis assumes that for each analysis, the missing data, in this case missing SGTF status, is independent from the outcome of interest, given the variables included in the models. This is a specific type of missing not at random (MNAR) assumption, as in particular it is allowed to depend on the underlying value of SGTF. We also performed an analysis of the complete cases using inverse probability weights[10] (IPW) to address the missing data on SGTF, under a missing at random (MAR) assumption."

To clarify this assumption, we have added a sentence to the Results section:

"All models were fitted twice, once using complete cases only, i.e. by simply excluding individuals with missing SGTF status, and once using inverse probability weighting (IPW), i.e. accounting for missingness by upweighting individuals whose characteristics—age, sex, IMD, ethnicity, residence type, NHS England region of residence and sampling week—are underrepresented among complete cases. **This analysis assumes that, holding these characteristics constant, whether an individual dies is independent of missingness in SGTF status[4].**"

Table 1 seems to have two columns that provide the same counts (although the percentages have different denominators). I'd suggest combining in some way.

Although there is redundant information in this table (now Extended Data Table 1), we feel that it is better to be explicit about the numerator and denominator for each percentage.

Table 2. Although a reader could work it out, percentages regarding missingness should be given. For example, the fact that ~74% of deaths have missing SGTF status (9,408/12790) is not evident.

We have added SGTF prevalence (%) and missingness (%) percentages to this table (now Extended Data Table 2).

We thank the reviewer for these helpful comments.